14.81K
Number of Sentences

439K
Total Number of Words

16.88K
Number of Distinct Words

POS
- ☐ Adjective
- ☐ Adverb
- ☐ Noun
- ☐ Other
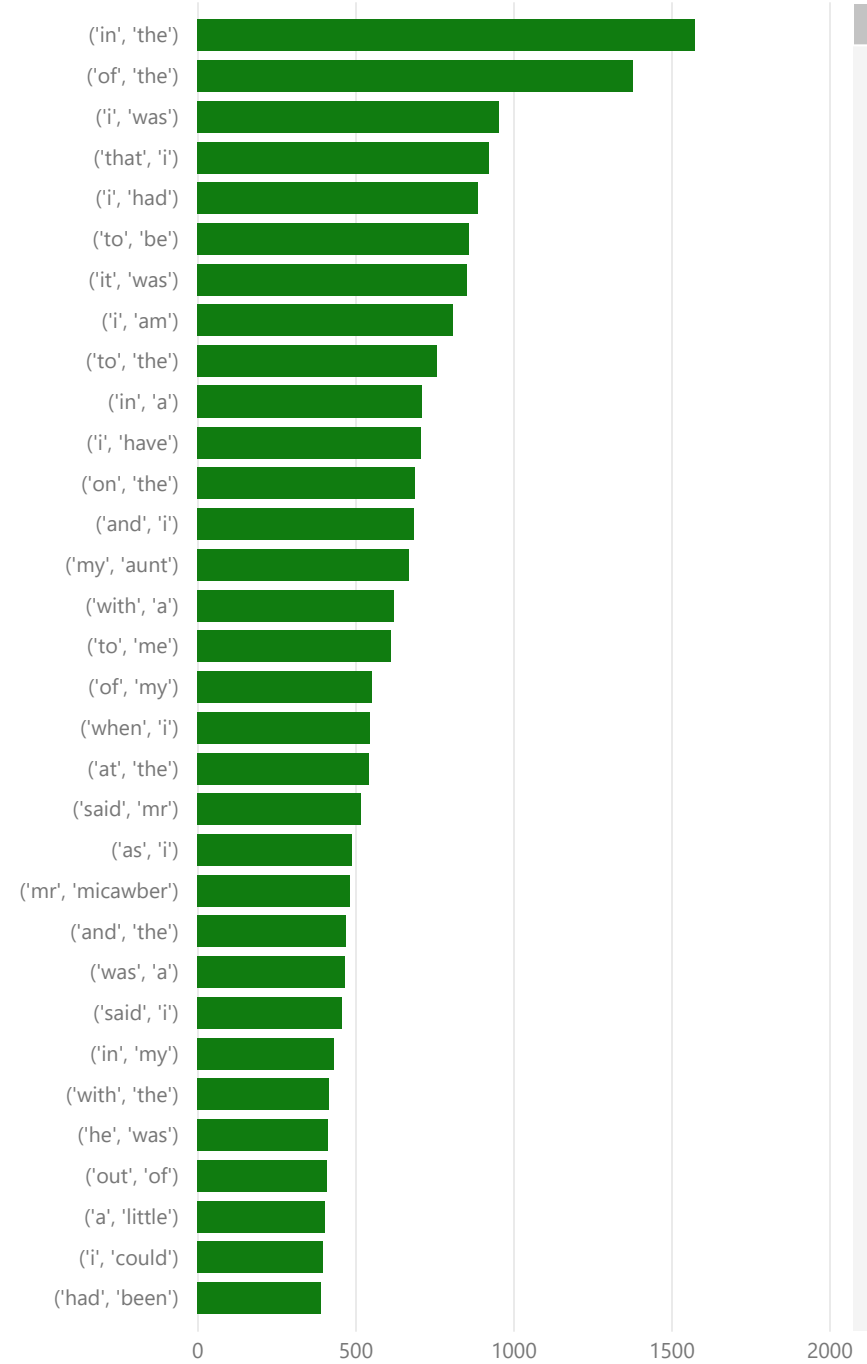- ☐ Verb

Sum of Count by POS
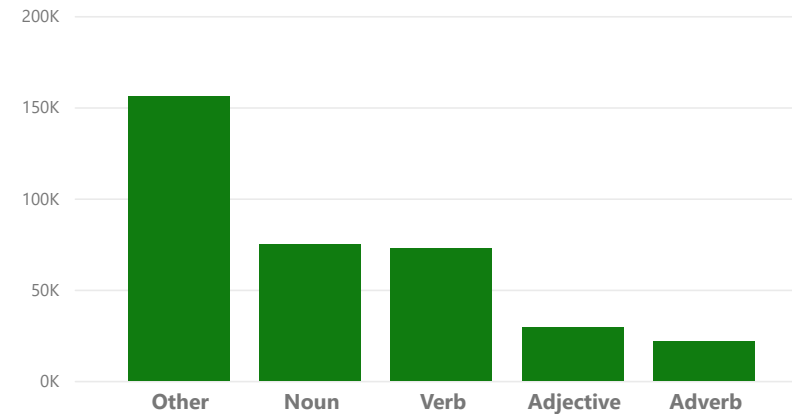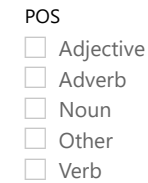
The word cloud prominently highlights frequently used words such as "little", "time", "know", "mother", "think", and "aunt", reflecting Dickens' emphasis on family relationships, personal growth, and the passage of time in this coming-of-age novel.

The metrics further reveal the novel's depth: with **439,000 total words, it exemplifies the substantial length typical of Victorian literature; 16,880 distinct words showcase Dickens' rich vocabulary;** and **14,810 sentences suggest a tendency toward relatively long sentence structures.**

Additionally, the distribution of parts of speech indicates that nouns and verbs dominate, while adjectives and adverbs are less frequent, reinforcing Dickens' focus on action and character development rather than excessive description.

Frequency by Bigram

Bigram	Frequency
('in', 'the')	~1560
('of', 'the')	~1380
('i', 'was')	~950
('that', 'i')	~910
('i', 'had')	~870
('to', 'be')	~850
('it', 'was')	~840
('i', 'am')	~790
('to', 'the')	~740
('in', 'a')	~700
('i', 'have')	~690
('on', 'the')	~680
('and', 'i')	~670
('my', 'aunt')	~650
('with', 'a')	~620
('to', 'me')	~600
('of', 'my')	~540
('when', 'i')	~540
('at', 'the')	~530
('said', 'mr')	~510
('as', 'i')	~480
('mr', 'micawber')	~470
('and', 'the')	~460
('was', 'a')	~450
('said', 'i')	~440
('in', 'my')	~420
('with', 'the')	~410
('he', 'was')	~400
('out', 'of')	~400
('a', 'little')	~390
('i', 'could')	~385
('had', 'been')	~380

Frequency by Trigram

| Trigram	Frequency
('said', 'my', 'aunt')	~215
('there', 'was', 'a')	~125
('i', 'could', 'not')	~122
('that', 'i', 'was')	~122
('out', 'of', 'the')	~120
('i', 'dont', 'know')	~115
('said', 'mr', 'peggotty')	~108
('it', 'was', 'a')	~105
('said', 'mr', 'micawb…')	~103
('i', 'am', 'sure')	~98
('if', 'i', 'had')	~95
('that', 'i', 'had')	~90
('i', 'had', 'been')	~85
('that', 'it', 'was')	~83
('as', 'if', 'he')	~80
('i', 'am', 'not')	~78
('i', 'have', 'been')	~70
('i', 'should', 'have')	~68
('a', 'good', 'deal')	~67
('my', 'aunt', 'and')	~65
('one', 'of', 'the')	~63
('i', 'can', 'not')	~62
('in', 'the', 'morning')	~60
('that', 'she', 'was')	~58
('i', 'have', 'no')	~57
('in', 'the', 'world')	~56
('that', 'he', 'was')	~55
('when', 'i', 'was')	~55
('said', 'mr', 'omer')	~54
('said', 'mrs', 'micawb…')	~53
('would', 'have', 'been')	~52
('that', 'i', 'could')	~50

These charts reveal Dickens' characteristic writing patterns, emphasizing his repeated use of certain phrase constructions. The prominence of **"said my aunt"** mentions the notable aunt characters, Betsey Trotwood.
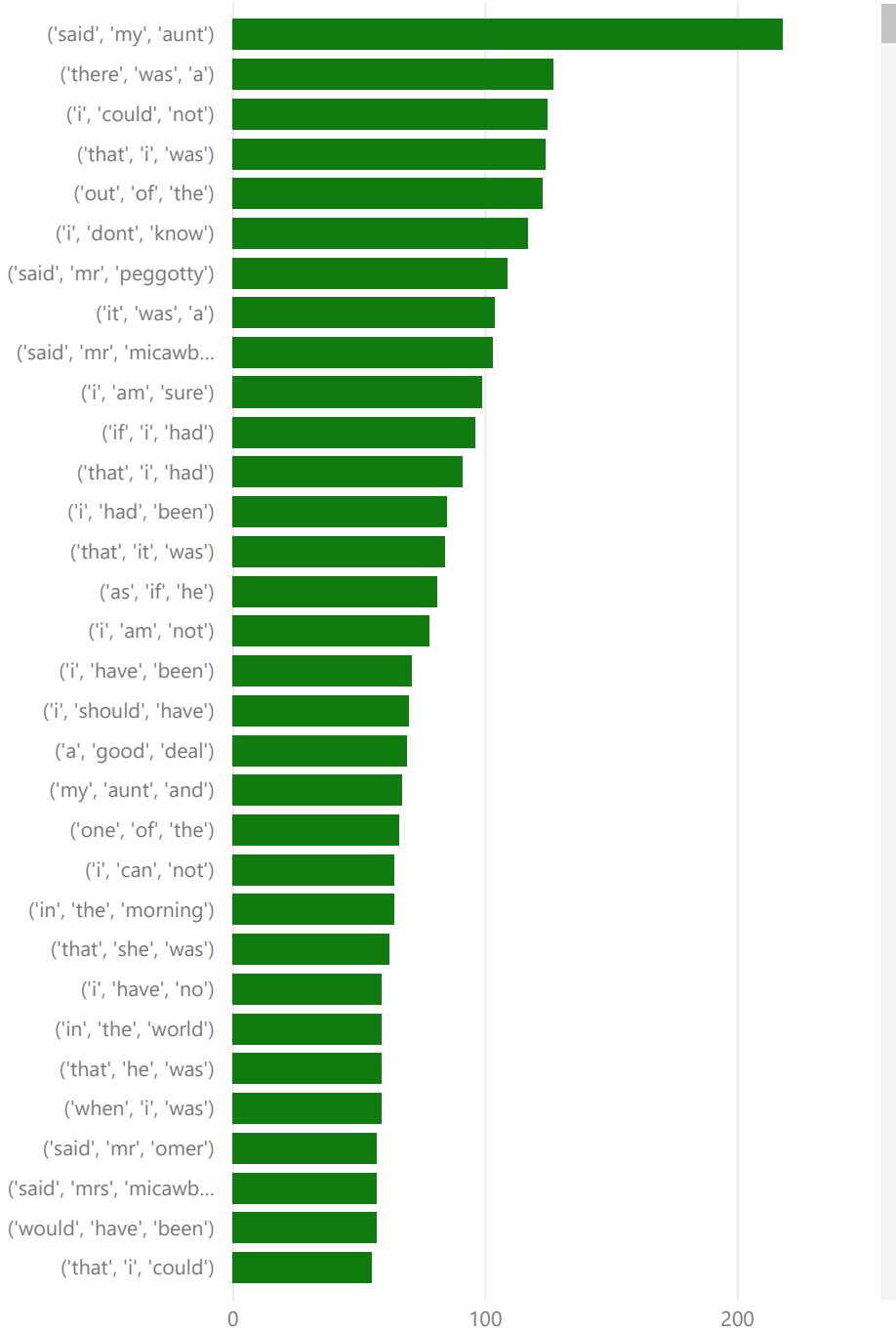
From_POS	Adjective	Adverb	Noun	Other	Verb
Verb	5808	8338	6075	41410	11035
Other	16575	6225	41991	57413	33975
Noun	1299	3850	9729	42101	17984
Adverb	3852	2663	652	7971	6788
Adjective	1858	850	16516	7283	2884

The scatter plot and the Trend Line indicates that **there is no consistent change in diversity across sentences.**

Sentence ID and POS Diversity
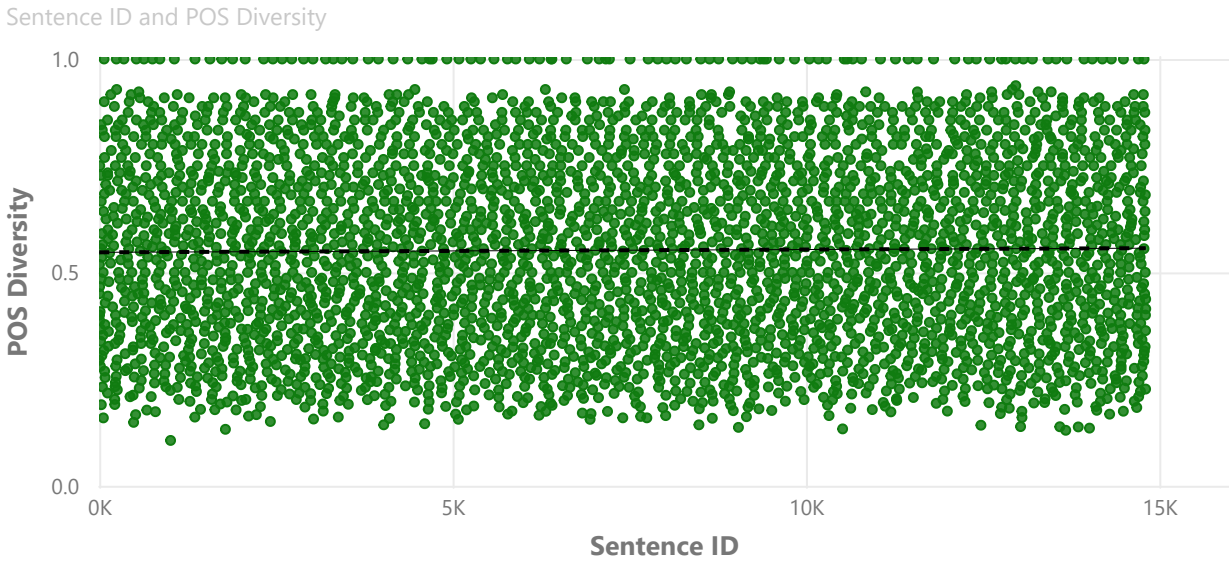
Average of Distances from various POS

● Average of Distance_from_last_Verb ● Average of Distance_from_last_Noun ● Average of Distance_from_last_Adjective